International Journal of English Literature and Social Sciences
Vol-9, Issue-4; Jul-Aug, 2024

Peer-Reviewed Journal
Journal Home Page Available: https://ijels.com/
Journal DOI: 10.22161/ijels

Multimodal Discourse Analysis of “Union Is Strength” from the Perspective of Visual Grammar

Huaming Cheng

Guangzhou College of Commerce, Guangzhou, China
kylechm@126.com

Received: 21 May 2024; Received in revised form: 29 Jun 2024; Accepted: 06 Jul 2024; Available online: 14 Jul, 2024

Abstract**—***As a typical multimodal discourse, a video can construct its meaning through various modalities, such as image, sound, action, and language. From the perspective of systemic functional grammar, taking the visual grammar of Kress & van Leeuwen as the theoretical framework, this paper analyzes the multimodal video “Union Is Strength” from the representational meaning, interactional meaning and compositional meaning, and discusses the meaning construction of the multimodality such as image and speech. The study has found that the video constructs union is strength in a variety of narrative ways, and union can realize the theme significance of the weak winning over the strong.*

Keywords**—*****compositional meaning, interactional meaning, multimodal discourse, representational meaning, visual grammar***

I. INTRODUCTION

The linguistic approach to discourse analysis is insufficient to thoroughly explore discourse, as the meaning of discourse incorporates non-linguistic factors. Conducting discourse analysis from a multimodal perspective that integrates images, sounds, languages, and actions can better interpret its meaning. The earliest analysis of multimodal discourse was conducted by the famous French linguist Barthes (1977), who explored the interaction between images and language in expressing meaning in his paper “Rhetoric of the image.” Based on Halliday’s metafunctional theory (ideational metafunction, interpersonal metafunction, and textual metafunction) in systemic functional grammar, Kress and van Leeuwen (2006) classified the meaning of images into representational meaning, interactional meaning, and compositional meaning, proposing a multimodal discourse analysis theory of visual grammar. Currently, research on visual grammar covers a wide range of topics, including the analysis of documentaries or promotional videos from the perspective of visual grammar, such as Wei and Li (2017), Wen (2019), Zhang *et al.* (2022), Cui and Zheng (2023), or the analysis of film works, such as Zheng (2016), Wang (2018), as well as film posters including Zhang (2013), Li (2020), and Shi (2022). There are also studies on the visual grammar of advertisements, such as Liu (2020), Ju (2020), on exploring picture books like Teng and Miao (2018), Chen and Chen (2019), or on analyzing news photos such as Dong and Wang (2020), and Wang (2021), etc. In general, visual grammar research involves various types of videos and images. This article attempts to analyze the construction of multimodal symbolic meaning in the video “Union Is Strength” from the perspective of visual grammar.

II. METHODOLOGY

This study adopts a combined quantitative and qualitative approach. By using the “Yuetu Video Frame Image Extractor v1.0,” multimodal corpora are collected through screenshot capturing of the 1-minute and 19-second-long video “Union Is Strength” at a rate of one screenshot per second, resulting in a total of 79 frames. After removing repeated images due to short capture intervals and blurry transitional frames, 33 frames are selected for analysis. The study examines the collected corpora from the three meanings of visual grammar to explore how images reproduce meaning through various symbols. The content studied in this article includes multimodal discourse encompassing various symbolic resources such as images, sounds, actions, and languages, which are texts realized through the encoding of multiple symbols.

III. RESULTS AND DISCUSSION

3.1 Representational Meaning

Based on the presence or absence of “vectors,” representational meaning is divided into “narrative representations” and “conceptual representations.” The hallmark of a narrative visual “proposition” is the presence of a vector: narrative structures always have one, conceptual structures never do. (Kress & van Leeuwen, 2006, p. 59) Among them, narrative representations include action processes, reactional processes, speech processes and mental processes, while conceptual representations without vectors can be further classified into classificational processes, analytical processes, and symbolic processes. In narrative representations, participants are connected through a vector, indicating that they are doing something to each other. Narrative representations are used to present ongoing actions and events, processes of change, and transient spatial arrangements, while conceptual representations refer to the class, structure, or meaning of participants; in other words, they refer more or less to stable and eternal essences. The hallmark of narrative representations is the existence of vectors, while conceptual representations do not have them.

3.1.1 Narrative Representations

Narrative representations are further subdivided into action processes, reactional processes, and speech and mental processes.

(1) Action Process

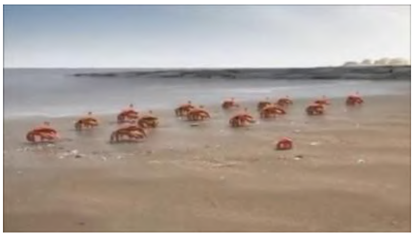

Fig. 1

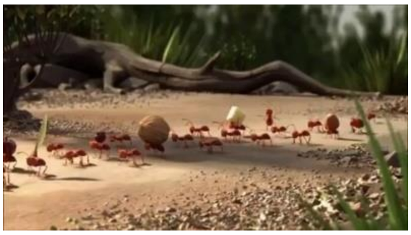

Fig. 2

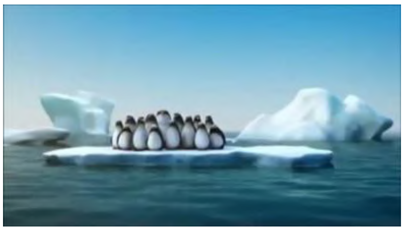

Fig. 3

In action process, the Actor is the participant from which the vector emanates, or which itself, in whole or in part, forms the vector (Kress & van Leeuwen, 2006, p. 63). The entire video of “Union Is Strength” falls into three segments. The first segment depicts crabs confronting a seagull; the second segment shows ants defending against an anteater; and the third segment portrays penguins fighting against a shark. In the action processes, the actor is the participant who generates the vector, or the actor itself (in whole or in part) forms the vector. At the beginning of the first segment, crabs crawl in a direction on the beach (Fig. 1); in the second segment, ants carry food along a path (Fig. 2); and in the third segment, penguins drift on the ice floe in the sea (Fig. 3). These are all action processes that contain vectors, exhibiting directionality and non-interactivity as there are no target objects.

Fig. 4

Fig. 5

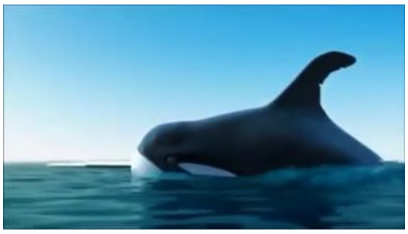

Fig. 6

When images or diagrams have only one participant, this participant is usually an actor. The resulting structure we call non-transactional. (Kress & van Leeuwen, 2006, p. 63) However, when the seagull starts diving towards the crabs (Fig. 4), the anteater forcefully sucks up the small ants (Fig. 5), and the shark swims underwater towards the penguins (Fig. 6), these are also action processes, displaying interactivity, with clear targets: the crabs, ants, and penguins become the objects, which are intended to be food.

(2) Reactional Process

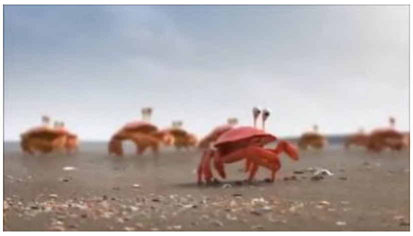

Fig. 7

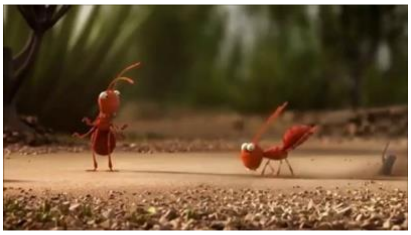

Fig. 8

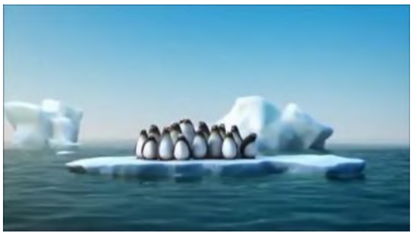

Fig. 9

When the vector is formed by an eyeline, by the direction of the glance of one or more of the represented participants, the process is reactional, and we will speak not of Actors, but of Reacters, and not of Goals, but of Phenomena. (Kress & van Leeuwen, 2006, p. 67) In the first segment, when a crab spots the seagull in the sky, it directs its two eyes towards it (Fig. 7). In this case, the crab is the reactor, and the seagull becomes the phenomenon. In the second segment, when another ant notices a small ant being sucked up by the anteater and directs its gaze towards the trapped ant, this ant becomes the reactor, while the small ant is the phenomenon (Fig. 8). In the third segment, when the penguin on the far right of the ice floe first discovers the shark, followed by all the penguins turning their attention towards the shark, the penguins are the reactors, and the shark becomes the phenomenon (Fig. 9). All of these scenarios are reactional processes, indicating that the confrontational battles of various underdogs have begun.

(3) Speech and Mental Process

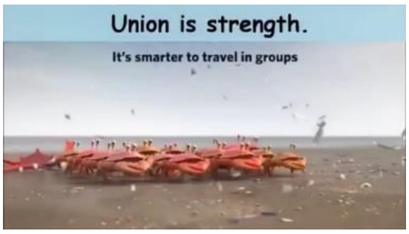

Fig. 10

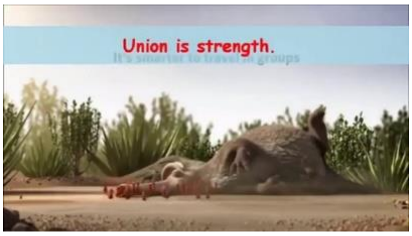

Fig. 11

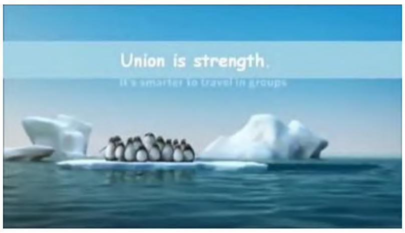

Fig. 12

A special kind of vector can be observed in comic strips: the oblique protrusions of the thought balloons and dialogue balloons that connect drawings of speakers or thinkers to their speech or thought. (Kress & van Leeuwen, 2006, p. 68) In this video, there are no thought balloons or dialogue balloons, but there is a small amount of sound. For instance, in the first segment, there is the surprised sound of the crab spotting the seagull, the seagull’s saliva-sucking sound upon seeing its delicacy (referring to the crab), then the clicking sound of the crab’s claws as it gathers all the crabs together, and finally, the pitiful screams of the seagull whose feathers are being clipped. The first segment ends with the crabs unifying and defeating the seagull (Fig. 10). Starting from the second segment, with the marching music of the ants, the sound of a small ant calling for help appears first, followed by another ant whistling to gather its companions. Under its command, the ants form a ball and finally block the anteater's nose to achieve victory, resulting in jubilant cheers from all the ants (Fig. 11). In the third segment, the penguin on the far right first utters a warning sound, prompting all the penguins to look towards the shark. They then command everyone to move to the left, tilt the ice floe, and finally let the shark crash into the iceberg to defeat it, ending with collective laughter (Fig. 12). At the end of each segment, two lines of text appear. The top line subtitle reads: "Union is strength," indicating the theme of the video, namely, unity is strength. When the subtitle at the bottom line appears, it is accompanied by a voiceover reading, "It's smarter to travel in groups," offering a suggestion that traveling in a group is a wiser choice.

3.1.2 Conceptual Representations

Conceptual representations can be subdivided into classificational processes, analytical processes, and symbolic processes.

(1) Classificational Process

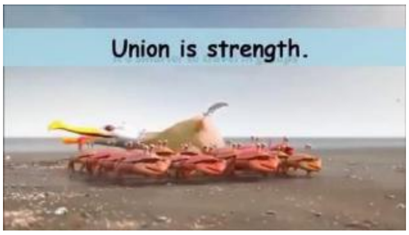

Fig. 13

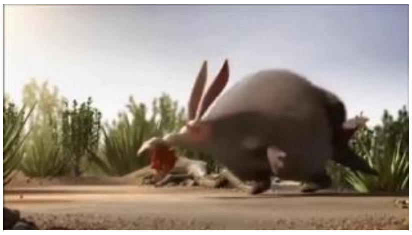

Fig. 14

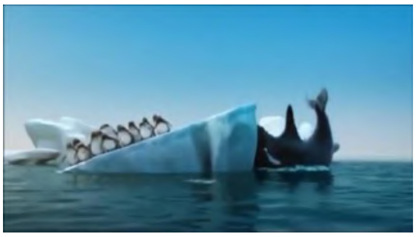

Fig. 15

Classificational processes relate participants to each other in terms of a 'kind of' relation, a taxonomy: at least one set of participants will play the role of Subordinates with respect to at least one other participant, the Superordinate. (Kress & van Leeuwen, 2006, p. 79) When two types of participants appear simultaneously, their relationship becomes apparent. For instance, the seagull (Fig. 13), the anteater (Fig. 14), and the shark (Fig. 15) are at the top of the food chain, while crabs, ants, and penguins occupy the lower levels. Such image incorporates the classificational process. As the top predators of the food chain, they ultimately end up being defeated by the participants at the lower levels, thus illustrating the importance of unity.

(2) Analytical Process

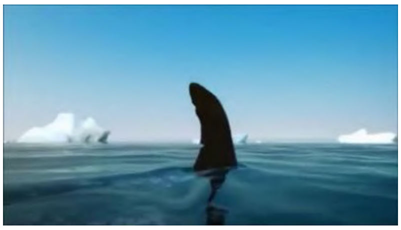

Fig. 16

Analytical processes relate participants in terms of a part–whole structure. They involve two kinds of participants: one Carrier (the whole) and any number of Possessive Attributes (the parts). (Kress & van Leeuwen, 2006, p. 87) As seen in the image above, the dorsal fin of the shark suggests that it is indeed a shark. Before it reveals its full presence, this part allows the penguins to associate it with a shark, thus causing them to become alert.

(3) Symbolic Process

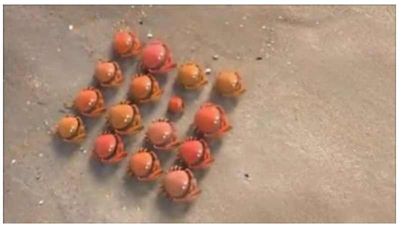

Fig. 17

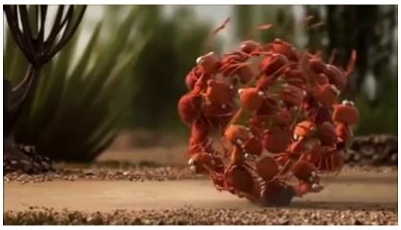

Fig. 18

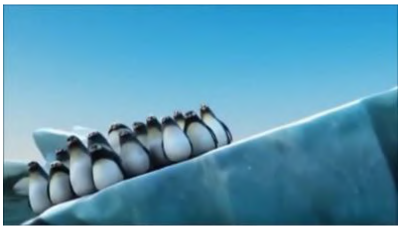

Fig. 19

Symbolic processes are about what a participant means or is. (Kress & van Leeuwen, 2006, p. 105) In the first segment, all the crabs gathered from all directions to form a square formation (Fig. 17); in the second segment, all the ants held hands and formed a sphere (Fig. 18); and in the third segment, all the penguins gathered on one side, lifting the ice floe to form a barrier (Fig. 19). These all symbolize unity, where they focus and unite into a force, which aligns with the thematic meaning of the video.

3.2 Interactional Meaning

The interactional meaning in visual grammar aims to explore the relationship between the image maker, the world presented in the image, and the viewer of the image, while expressing the attitude that the viewer should hold towards the represented object. Generally speaking, the expression of interactional meaning includes four dimensions: contact, social distance, perspective, and modality.

(1) Contact

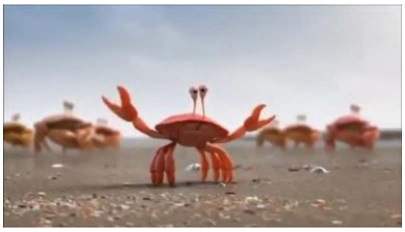

Fig. 20

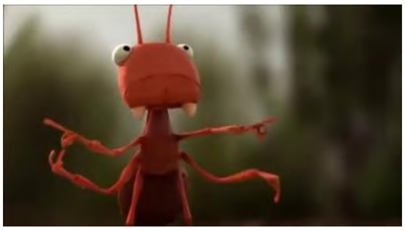

Fig. 21

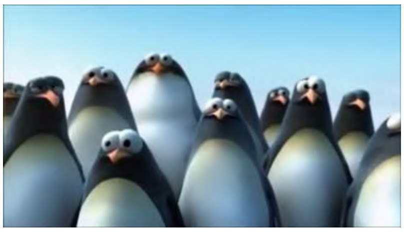

Fig. 22

When represented participants look at the viewer, vectors, formed by participants' eyelines, connect the participants with the viewer. Contact is established, even if it is only on an imaginary level. (Kress & van Leeuwen, 2006, p. 117) "Contact" refers to an imaginary contact relationship established between the participants in the image and the viewer through the direction of gaze, which can be divided into "demand" and "offer." "Demand" images refer to those where the participants in the image have eye contact with the viewer, seeking information or something else. "Offer" images generally do not involve eye contact and only provide information to the reader. When the represented participant looks at the viewer, a vector formed by the participant's gaze connects the participant and the viewer. Contact is thus established, even if it is only on an imaginary level. The producer uses the image to do something to the audience. It is for this reason that we refer to such images as "demand." All images that do not include human or anthropomorphic participants looking directly at the viewer are classified as "offer." In some cases, such as television news reports and posed magazine photos, "demand" images are preferred: these situations require a connection between the viewer and the authorities, celebrities, and role models they depict. In other cases, such as feature films, TV dramas, and scientific illustrations, "offer" is preferred. When the small crab confidently calls for the crabs to gather facing the viewer (Fig. 20), when the calm and commanding ant faces the viewer (Fig. 21), and when all the penguins' gaze meets the viewer (Fig. 22), these images are classified as "demand" because the participants in the video have eye contact with the viewer, establishing a connection between them. Other images without eye contact are classified as "offer," providing specific information.

(2) Social Distance

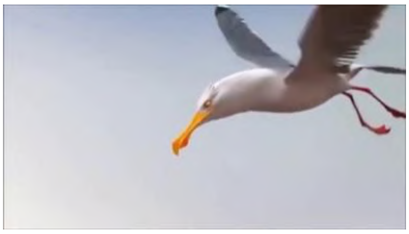

Fig. 23

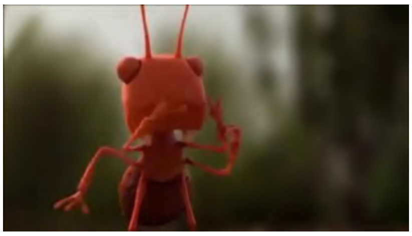

Fig. 24

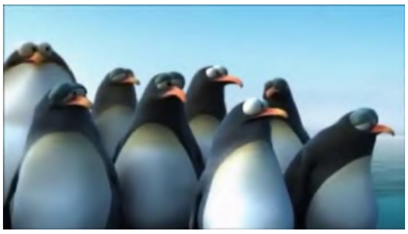

Fig. 25

The close shot shows head and shoulders of the subject. The medium close shot cuts off the subject approximately at the waist. In the long shot the human figure occupies about half the height of the frame. (Kress & van Leeuwen, 2006, p. 124) The content of the three clips is primarily shot in long shots. In the first clip, there is a medium shot only when the seagull dives down (Fig. 23). In the second clip, there is a medium shot only when the commanding ant is filmed (Fig. 24). In the third clip, there is a medium shot when filming the penguins (Fig. 25), and the rest are long shots. In real life, social relations determine the distance that should be maintained between people. Kress & van Leeuwen believe that the framing size of an image can also reflect the closeness or distance between the viewer and the participants in the image. For example, a close-up shot represents an intimate distance, a full-body shot of a person in the frame represents a social close distance, a full-body shot of a person in the frame with space surrounding it represents a social long distance, and a distance of no less than 3 to 4 people is referred to as a public distance.

(3) Perspective

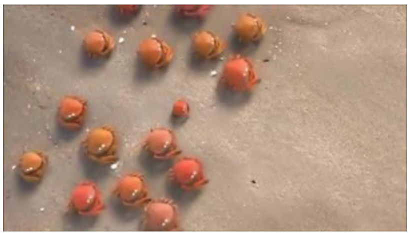

Fig. 26

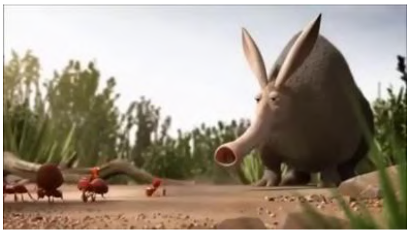

Fig. 27

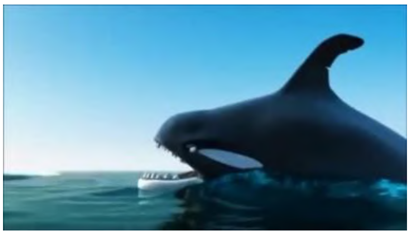

Fig. 28

Producing an image involves not only the choice between "offer" and "demand" and the selection of a certain size of frame, but also the selection of an angle, a "point of view." (Kress & van Leeuwen, 2006, p. 129) Making images is not only a choice between "offer" and "demand," but also requires a certain size of the frame, and also requires a choice of angles and perspective, which means it can express subjective attitudes towards the participants being reproduced, whether they are human or other things. Most shots in the whole video are eye-level. Compared with the seagull and the crabs, the anteater and the ants, the shark and the penguins, the former are relatively large in size or volume, and in a strong position. In the first segment, the perspective is overlooking from the seagull's perspective (Fig. 26), so the seagull is in a strong position. In the second segment, the ants face the larger anteater, facing it from a perspective of looking up (Fig. 27), so the ant is in a weak position. In terms of size, the shark is large and in a dominant position (Fig. 28).

(4) Modality

Modality refers to the credibility or authenticity of people's statements about the world they are concerned about, and its manifestation of resources is relatively rich in images. Similar to systemic functional linguistics, modality is categorized into three levels: high, medium, and low. This involves three scales, discussing the role of color as a naturalistic modal marker: (1) Color saturation, ranging from full color saturation to achromatic, i.e., becoming black and white. (2) Color differentiation, ranging from a maximum range of diverse colors to a monochromatic range. (3) Colour modulation, a scale running from fully modulated colour, with, for example, the use of many different shades of red, to plain, unmodulated colour. (Kress & van Leeuwen, 2006, p. 160) As an animated image, this video has maximum color saturation, making it appear "hyperrealistic."

3.3 Compositional Meaning

Compositional meaning corresponds to the textual meaning in functional grammar, referring to how an image integrates its representational meaning and interactional meaning to form a meaningful whole. Compositional meaning comprises three resources: information value, salience, and framing.

(1) Information Value

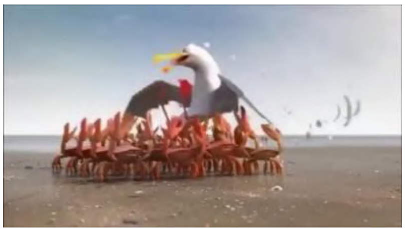

Fig. 29

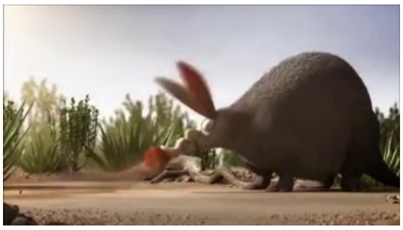

Fig. 30

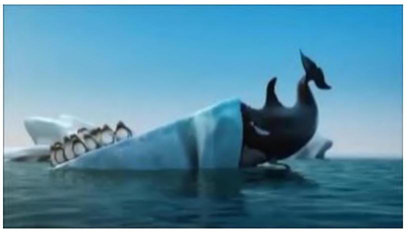

Fig. 31

The placement of elements endows them with the specific informational values attached to the various 'zones' of the image: left and right, top and bottom, centre and margin. (Kress & van Leeuwen, 2006, p. 177) Elements placed on the left represent old information, while elements placed on the right represent new information. Things placed at the top are considered ideal, while those at the bottom are seen as reality. For example, in the case of top and bottom, the seagull is diving down from the sky, resulting in the seagull being clipped (Fig. 29). As for left and right, the ants are on the left and the anteater is on the right (Fig. 30); the penguins are on the left and the shark is on the right (Fig. 31). The new information appearing on the right, which seems to represent the stronger party, ends in failure.

(2) Salience

The elements are made to attract the viewer's attention to different degrees, as realized by such factors as placement in the foreground or background, relative size, contrasts in tonal value (or colour), differences in sharpness, etc. (Kress & van Leeuwen, 2006, p. 177) As seen in the three images mentioned above, there are differences in size and contrasts in strength, but the outcomes are all failures: one has its feathers clipped, one is blocked by a nostril, and the other crashes into an iceberg. These three endings indicate that the power of unity can enable the weak to defeat the strong.

(3) Framing

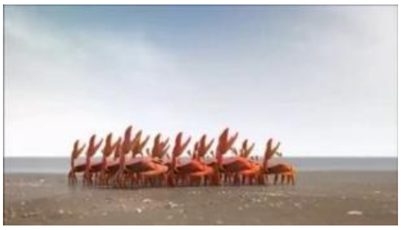

Fig. 32

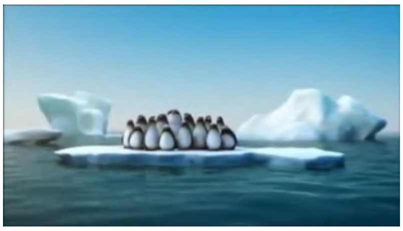

Fig. 33

The presence or absence of framing devices disconnects or connects elements of the image, signifying that they belong or do not belong together in some sense. (Kress & van Leeuwen, 2006, p. 177) There is no dedicated frame in the video, but in fact, in the first segment, the crabs' claws reach up to the sky, confronting the diving seagull, with the ground and the sky forming a dividing line (Fig. 32). Similarly, in the third segment, the penguins stand on an ice floe above the sea, with the ice floe and the sea forming a line, creating a contrast between the penguin on the sea surface and the shark in the sea (Fig. 33).

IV. CONCLUSION

This article uses visual grammar by Kress and van Leeuwen as a theoretical framework to analyze the representational meaning, interactional meaning, and compositional meaning of multimodal symbols in the video "Union Is Strength" from various modalities such as images, sounds, languages, and actions. Its purpose is to reveal the interaction between image symbols and linguistic symbols in constructing video meaning. The research shows that the video constructs the power of unity through multiple narrative methods, demonstrating that the weak can defeat the strong when united. This study can expand the application range of visual grammar theory and verify its operability and practicability. However, as the study mainly uses qualitative research methods, the research results inevitably suffer from personal subjective speculation and possess a certain degree of subjectivity. It is hoped that future research on multimodal discourse will tend to be more quantitative.

REFERENCES

[1] Barthes, R. (1977). Rhetoric of the image. *Image-Music-Text*. Sel. and Trans. Stephen Heath. Hill and Wang, 32-51.

[2] Chen, D., & Chen, Z. (2019). A comparative of Chinese and English children's picture books in narrative construction based on new visual grammar. *Journal of Xi'an International Studies University, 27*(04), 36-41.

[3] Cui, W. Y. & Zheng, L. (2023). A Multimodal Discourse Analysis of *Planet Earth II* from the Perspective of Visual Grammar. *Journal of Humanities, Arts and Social Science, 7*(12), 2455-2459.

[4] Dong, Y., & Wang, X. (2020). Analysis of Visual Framework Construction and Visual Grammar in News Images Reporting Terrorist Attacks: Taking People's Daily as an Example. *China Publishing Journal,* (06), 32-36.

[5] Ju, X. (2020). Multimodal Discourse Analysis of Public Service Advertisements from the Perspective of Visual Grammar: Taking the Public Service Advertisement "No Hospital, No Hope" as an Example. *Today's Mass Media, 28*(02), 88-90.

[6] Kress, G. & van Leeuwen, T. (2006). *Reading Images: The*

Grammar of Visual Design. Routledge.

[7] Li, S. (2021). Interpretation of Interactive and Compositional Meanings of Film Posters from the Perspective of Visual Grammar: Taking the Poster of "Parasite" as an Example. *Popular Culture and Arts*, (08), 151-152.

[8] Liu, D. (2020). Multimodal Metaphor Construction of Vertical Screen Microfilm Advertisements from the View of Visual Grammar—Take Huawei's Advertisement Wukong as an Example. *Journal of Huaqiao University (Philosophy & Social Sciences)*, (01), 154-160.

[9] Shi, Y. (2022). Multimodal Discourse Analysis of the Movie Poster of "Chinese Doctors" from the Perspective of Visual Grammar. *Jingu Creative Literature*, (10), 87-89.

[10] Teng, D., & Miao, X. (2018). Meaning Construction of Multi-modal Metaphors in the Picture Book Discourse from the Grammar of Visual Design. *Foreign Language Research*, (05), 53-59.

[11] Wang, M. (2018). The Construction of National Image in "Wolf Warriors II" from the Perspective of Visual Grammar. *Journalism Lover*, (02), 81-84.

[12] Wang, N. (2021). The Construction of Image Meaning in News Photography of Public Health Emergencies Based on Visual Grammar: One of the Studies on News Images of the "Fight Against the Epidemic". *Journalism Lover,* (01), 83-86.

[13] Wei, B., & Li, C. (2017). Multimodal Analysis on "the Belt and Road" Propaganda Film from the View of Visual Grammar. *Journal of Harbin University, 38*(01), 130-135.

[14] Wen, W. (2019). Multimodal analysis of the English documentary "One Belt, One Road" from the perspective of visual grammar. *Journal of Hunan University of Science and Engineering, 40*(05), 128-129.

[15] Zhang, J. (2013). Interpretation of the Imagery of the Poster of "Life of Pi" from the Perspective of Visual Grammar. *Movie Review*, 16, 64-65.

[16] Zhang, W., Han, X., Wang, F., Liu, Q., Lin, X., & Yao, X. (2022). Research on the Construction of National Image in the Promotional Video for the Beijing Winter Olympic Games from the Perspective of Visual Grammar. *Comparative Study of Cultural Innovation, 6*(29), 64-68.

[17] Zheng, X. (2016). Analysis of the interactional meaning of visual grammar in the film "The Pursuit of Happyness". *Education and Teaching Forum*, (29), 108-109.